62 (2017) APPLICATIONS OF MATHEMATICS No. 6, 537–559

PRECONDITIONING OF TWO-BY-TWO BLOCK MATRIX SYSTEMS WITH SQUARE MATRIX BLOCKS, WITH APPLICATIONS

Owe Axelsson, Ostrava

Received August 16, 2017. First published December 4, 2017.

This paper is dedicated to the memory of Ivo Marek,
who recently unexpectedly deceased.
The author is particularly thankful for his long lasting friendship with Ivo Marek,
which also resulted in establishing important contacts with other
excellent numerical analysts in the Czech Republic.

Abstract. Two-by-two block matrices of special form with square matrix blocks arise in important applications, such as in optimal control of partial differential equations and in high order time integration methods.

Two solution methods involving very efficient preconditioned matrices, one based on a Schur complement reduction of the given system and one based on a transformation matrix with a perturbation of one of the given matrix blocks are presented. The first method involves an additional inner solution with the pivot matrix block but gives a very tight condition number bound when applied for a time integration method. The second method does not involve this matrix block but only inner solutions with a linear combination of the pivot block and the off-diagonal matrix blocks.

Both the methods give small condition number bounds that hold uniformly in all parameters involved in the problem, i.e. are fully robust. The paper presents shorter proofs, extended and new results compared to earlier publications.

Keywords: preconditioning; Schur complement; transformation; optimal control; implicit time integration

MSC 2010: 65F08

This work was supported by The Ministry of Education, Youth and Sports of the Czech Republic from the National Programme of Sustainability (NPU II), project "IT4Innovations excellence in science—LQ1602".

DOI: 10.21136/AM.2017.0222-17

1. Introduction

Two by two block matrix systems with square matrix blocks arise in several important applications such as when solving certain optimal control problems [15], [3], [4], when solving complex valued systems in real arithmetics [6] and, for instance, in the two point Radau time-integration method to solve systems of ordinary differential equations [1], [2]. In such problems, the system matrix can be written in the form

$$\mathcal{A} = \begin{bmatrix} D_1 & -L_2 \\ L_1 & D_2 \end{bmatrix},$$

where D_i, $i = 1, 2$, have equal order, are often symmetric positive definite (spd) and $L_1 + L_2$ is symmetric and positive semidefinite. In some problems $L_2 = cL_1$ or $L_2 = cL_1^{\mathrm{T}}$, for some positive scalar c. A matrix factorization of $\mathcal{A}$ gives

$$\mathcal{A} = \begin{bmatrix} D_1 & 0 \\ L_1 & S \end{bmatrix} \begin{bmatrix} I & -D_1^{-1}L_2 \\ 0 & I \end{bmatrix},$$

where $S = D_2 + L_1 D_1^{-1} L_2$. An application of this factorization requires two solutions with matrix D_1 and one with the Schur complement matrix S. The latter often implies a heavy computational cost. However, the systems can be solved by iteration so it is not required to form S explicitly. As we shall see, in some problems one can construct a very efficient preconditioner to S. To handle the cases where this is not possible, to reduce the cost we consider also some special transformed forms of the preconditioning method with no need to use S. This is particularly important when we solve optimal control problems with a state equation that involves its own constraint, such as the Stokes equation. The paper presents in a uniform way new, shorter and more generally applicable methods both for their derivation, implementation and eigenvalue analyses and surveys some important applications.

An efficient implementation of the method is derived where each iteration involves only two matrix-vector products besides the solution of two systems with a matrix that is a linear combination of the two block row matrices. This is even less than the matrix-vector multiplications involved in a matrix-vector multiplication with the given matrix.

For reasons of comparison, we present also another method, the preconditioned modified Hermitian and Skew-Hermitian splitting (PMHSS) iteration method, see [9], [10], that has been presented there to solve a more special class of problems than we are dealing with. This method involves also inner systems with a matrix that is a linear combination of the block matrices. We present new and much shorter proofs of the eigenvalue bounds for this method. It can be seen that it is not competitive with our method.

The paper is composed as follows. In Section 2 we consider the Schur complement based approach. Then in Section 3 we consider the transformed form of preconditioner, where we first assume that the block diagonal matrices are positive definite. This assumption is then relaxed in Section 4, where we consider a Radau time integration method for the time-dependent Stokes equation. In Section 5 the short and more general derivation of eigenvalue bounds for the PMHSS method is given. Section 6 contains some comments on the use of a generalized conjugate gradient method. The paper ends with some concluding remarks.

2. A Schur complement based solution method

2.1. A preconditioner for the Schur complement matrix. A linear block matrix system

$$\mathcal{A}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix},$$

can be solved via the Schur complement by first eliminating x_1, $x_1 = D_1^{-1}(f_1 + L_2 x_2)$ and substituting it in the second equation, $D_2 x_2 + L_1 x_1 = f_2$ to form the Schur complement residual,

$$r_2 = S x_2 - (f_2 - L_1 D_1^{-1} f_1) = D_2 x_2 - f_2 + L_1 D_1^{-1}(L_2 x_2 + f_1).$$

This is normally solved by iteration, that is, given x_2 then solve (approximately) $S(\delta x_2) = r_2$, and let $x_2 := x_2 + \delta x_2$. Besides some matrix-vector multiplications the evaluation of an action of S costs one solution of a system with matrix D_1 per iteration. To get an acceptable solution cost, thereby one must use an efficient preconditioner to S. In some applications, such as when solving complex valued systems in real arithmetics and for optimal control problems we have $D_1 = D_2 = D$, an spd matrix and $L_1 = aL$, $L_2 = bL^{\mathrm{T}}$, where a and b have the same sign.

In this case

$$S = D + abLD^{-1}L^{\mathrm{T}}. \tag{2.1}$$

As a preconditioner we then use

$$S_0 = (D + \gamma L)D^{-1}(D + \gamma L^{\mathrm{T}}), \tag{2.2}$$

where γ is a parameter to be chosen. We need to compute eigenvalue bounds that hold for the preconditioned matrix $S_0^{-1}S$.

In the next proposition we generalize this problem to make it applicable also for more general matrices, such as the matrix appearing in Section 2.3.

Proposition 2.1. *Let A, of order $n \times n$, be spd and assume that $B + B^{\mathrm{T}}$ is positive semidefinite. Let $S = A + c(B + B^{\mathrm{T}}) + d^2 B A^{-1} B^{\mathrm{T}}$, where $0 \leqslant c \leqslant d$, $d > 0$ and let $S_0 = (A + \gamma B)A^{-1}(A + \gamma B^{\mathrm{T}})$ be a preconditioner to the S, where $\gamma = d$. Then the eigenvalues λ of $S_0^{-1}S$ are contained in the interval $\delta = \frac{1}{2}(1 + c/d) \leqslant \lambda \leqslant 1$.*

P r o o f. Consider the generalized eigenvalue problem, $\lambda S_0 x = Sx$, $x \neq 0$. A congruence transformation with $A^{-1/2}$ of both matrices gives

$$\lambda \widetilde{S}_0 \tilde{x} = \widetilde{S}\tilde{x},$$

where $\widetilde{S}_0 = A^{-1/2} S_0 A^{-1/2}$, $\widetilde{S} = A^{-1/2} S A^{-1/2}$, and $\tilde{x} = A^{1/2} x$. Hence,

$$\begin{aligned}\lambda &= \frac{\tilde{x}^{\mathrm{T}} \widetilde{S} \tilde{x}}{\tilde{x}^{\mathrm{T}} \widetilde{S}_0 \tilde{x}} = \frac{\tilde{x}^{\mathrm{T}}(I + c(\widetilde{B} + \widetilde{B}^{\mathrm{T}}) + d^2 \widetilde{B}\widetilde{B}^{\mathrm{T}})\tilde{x}}{\tilde{x}^{\mathrm{T}}(I + \gamma(\widetilde{B} + \widetilde{B}^{T}) + \gamma^2 \widetilde{B}\widetilde{B}^{\mathrm{T}})\tilde{x}} \\ &= 1 - \frac{(\gamma - c)\tilde{x}^{\mathrm{T}}(\widetilde{B} + \widetilde{B}^{\mathrm{T}})\tilde{x}}{\tilde{x}^{\mathrm{T}}(I + \gamma(\widetilde{B} + \widetilde{B}^{T}) + \gamma^2 \widetilde{B}\widetilde{B}^{\mathrm{T}})\tilde{x}} \leqslant 1,\end{aligned}$$

where $\widetilde{B} = A^{-1/2} B A^{-1/2}$. Since $\tilde{x}^{\mathrm{T}}(I - \gamma \widetilde{B})(I - \gamma \widetilde{B}^{\mathrm{T}})\tilde{x} \geqslant 0$, it follows that

$$\gamma \tilde{x}^{\mathrm{T}}(\widetilde{B} + \widetilde{B}^{\mathrm{T}})\tilde{x} \leqslant \tilde{x}^{\mathrm{T}}\tilde{x} + \gamma^2 \tilde{x}^{\mathrm{T}} \widetilde{B}\widetilde{B}^{\mathrm{T}} \tilde{x}.$$

Therefore,

$$\lambda \geqslant 1 - \frac{(\gamma - c)\tilde{x}^{\mathrm{T}}(\widetilde{B} + \widetilde{B}^{\mathrm{T}})\tilde{x}}{2\gamma \tilde{x}^{\mathrm{T}}(\widetilde{B} + \widetilde{B}^{T})\tilde{x}} = \frac{1}{2}\Big(1 + \frac{c}{\gamma}\Big).$$

□

Applying this proposition for the matrix in (2.1) yields $c = 0$, $d^2 = ab$, so for the eigenvalues of $S_0^{-1}S$, where S_0 is defined in (2.2) with $\gamma^2 = ab$, we obtain $\frac{1}{2} \leqslant \lambda(S_0^{-1}S) \leqslant 1$.

It follows that the condition number of $S_0^{-1}S$ is bounded by 2. Consider now two applications.

2.2. An optimal control problem. In optimal control problems for PDEs one normally introduces a Lagrange multiplier. Then the system of first order necessary optimality conditions for the corresponding Lagrangian functional includes as a subsystem a matrix of the form discussed in this paper.

We consider then the problem of finding the optimal solution u, f that minimizes the functional

$$J(u, f) = \tfrac{1}{2}\|u - u_d\|^2 + \tfrac{1}{2}\beta \|f\|^2,$$

where u_d is the desired (target) solution, $f \in L^2(\Omega)$ is a distributed control function and u satisfies the differential equation

$$\mathcal{L}u = f \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \Gamma = \partial\Omega.$$

Here Ω is a bounded simply connected domain in $\mathbb{R}^n$, $n = 1, 2$ or 3. Further, $\beta > 0$ is a regularization parameter. To illustrate the problem, we take

$$\mathcal{L}u = -\nabla(\nu\nabla u) + \sigma u,$$

where $\nu > 0$, $\sigma \geqslant 0$.

The differential equation constraint is implemented via a Lagrange multiplier (v). To strengthen the inf-sup condition, i.e. the saddle point structure of the problem, we use an augmented formulation by adding the term $-\frac{1}{2}\alpha\|v\|^2$, $\alpha \geqslant 0$ to the Lagrangian functional, which therefore takes the form

$$\mathcal{L}(u, v, f) = J(u, f) + \int_\Omega v(\mathcal{L}u - f)\,\mathrm{d}x - \tfrac{1}{2}\alpha\|v\|^2.$$

Using partial integration and the homogeneous boundary conditions, $u = 0$ and $v = 0$ on Γ, we can rewrite the Lagrange multiplier term in a symmetric bilinear form as

$$\int_\Omega v(\mathcal{L}u - f)\,\mathrm{d}x = \int_\Omega (\nu\nabla u \cdot \nabla v + \sigma uv - fv)\,\mathrm{d}x.$$

This leads to an adjoint differential equation for the Lagrange multiplier $-\Delta v + \sigma v = f$ in Ω, $v = 0$ on Γ.

Applying the first order necessary conditions, which are also sufficient for the existence of a solution, using equal finite element approximations in $H_0^1(\Omega)$ for u and v and a finite element approximation in $L^2(\Omega)$ for f, we get a linear system,

$$\text{(2.3)} \qquad \begin{cases} \dfrac{\partial\mathcal{L}}{\partial\underline{u}} = 0\colon\ M(\underline{u} - \underline{u}_d) + K\underline{v} = 0, \\ \dfrac{\partial\mathcal{L}}{\partial\underline{v}} = 0\colon\ K\underline{u} - \alpha M\underline{v} - \widetilde{M}_0^{\mathrm{T}}\underline{f} = 0, \\ \dfrac{\partial\mathcal{L}}{\partial\underline{f}} = 0\colon\ \beta M_0\underline{f} - \widetilde{M}_0\underline{v} = 0. \end{cases}$$

Here M is the mass matrix $[\int_\Omega \varphi_i\varphi_j]$, corresponding to basis functions $\{\varphi_i\}$ for u and v, K is the symmetric stiffness matrix corresponding to the operator $\mathcal{L}$ and M_0 is the mass matrix corresponding to basis functions $\{\Psi_i\}$ for $\underline{f} \in L_2(\Omega)$ and $\widetilde{M}_0 = \int_\Omega \Psi_i\Phi_j$. Note that $\widetilde{M}_0$ is a rectangular matrix.

In this problem we may eliminate f to form the system

$$\begin{bmatrix} M & K \\ K & -\alpha M - \frac{1}{\beta}\hat{M}_0 \end{bmatrix} \begin{bmatrix} \underline{u} \\ \underline{v} \end{bmatrix} = \begin{bmatrix} M\underline{u}_d \\ 0 \end{bmatrix},$$

where $\hat{M}_0 = \widetilde{M}_0^{\mathrm{T}} M_0^{-1} \widetilde{M}_0$.

This system can be solved via the reduced Schur complement equation,

$$Sv = Mu_d,$$

where

$$S = M_1 + KM^{-1}K, \quad M_1 = \alpha M + \frac{1}{\beta}\hat{M}_0.$$

To form a preconditioner to S we first approximate M_1 by $\tilde{\alpha}M$, $\tilde{\alpha} > 0$. If the basis functions $\Psi_i = \varphi_i$, then $M_1 = (\alpha + 1/\beta)M$, so $\tilde{\alpha} = (\alpha + 1/\beta)$. We let now

$$S_0 = \big(\sqrt{\tilde{\alpha}}M + \gamma K\big)M^{-1}\big(\sqrt{\tilde{\alpha}}M + \gamma K\big)$$

be the preconditioner to the Schur complement matrix

$$S = \tilde{\alpha}M + KM^{-1}K.$$

Taking $\gamma = 1$, it follows from Proposition 2.1 that we get a condition number bound $\kappa(S_0^{-1}S) \leqslant 2$. The method in Section 3 gives an alternative choice where the inner system with matrix M, needed to evaluate actions of the Schur complement matrix, is avoided.

2.3. An implicit time-integration method. There exist high order time-integration methods that are strongly A-stable, such as those based on Radau integration, see e.g. [12], [1].

Consider here a two-point Radau integration method to solve a system of ordinary differential equations,

$$M\frac{\mathrm{d}x}{\mathrm{d}t} + \hat{A}x(t) = f(t), \quad t > 0, \quad x(0) = x_0,$$

where M is spd, frequently a mass matrix, and $\hat{A} + \hat{A}^{\mathrm{T}}$ is assumed to be positive semidefinite. The global time-integration error in this method is $O(\tau^3)$, i.e. of a higher order than in the familiar Crank-Nicolson method. Furthermore, the Radau time-integration method is strongly, i.e. asymptotically, stable.

Here the matrix systems to be solved at each time t with time step τ takes the form (see, e.g. [1], [2])

$$\begin{bmatrix} M + \frac{5}{12}A & -\frac{1}{12}A \\ \frac{9}{12}A & M + \frac{3}{12}A \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} Mx_0 + \frac{1}{12}\tau(5f_1 - f_2) \\ Mx_0 + \frac{1}{4}\tau(3f_1 + f_2) \end{bmatrix},$$

where $A = \tau\hat{A}$. Multiplying by $\begin{bmatrix} M^{-1} & 0 \\ 0 & M^{-1} \end{bmatrix}$, eliminating variable x_1, using commutativity between matrix products and multiplying back with M, we get a reduced block matrix system

$$Sx_2 = \Big[M + \frac{2}{3}A + \frac{1}{6}AM^{-1}A\Big]x_2 = \Big(M - \frac{1}{3}A\Big)x_0 + \frac{3}{4}\tau f_1 + \tau\Big(\frac{1}{4}M + \frac{1}{6}A\Big)M^{-1}f_2$$

to be solved. The reduced matrix S can here be preconditioned with

$$S_0 = (M + \gamma A)M^{-1}(M + \gamma A) = M + 2\gamma A + \gamma^2 AM^{-1}A.$$

Proposition 2.2. *Let $\gamma = 1/\sqrt{6}$. The eigenvalues of the preconditioned matrix $S_0^{-1}S$ are contained in the interval $[\delta, 1]$, where $\delta = \frac{1}{2}\big(1 + \sqrt{\frac{2}{3}}\big)$.*

P r o o f. This follows from Proposition 2.1 with $c = \frac{1}{3}$ and $d^2 = \frac{1}{6}$, so the lower bound becomes $\delta = \frac{1}{2}(1 + \frac{1}{3}\sqrt{6}) = \frac{1}{2}(1 + \sqrt{\frac{2}{3}})$. □

It is seen that the eigenvalues of $\widetilde{S}_0^{-1}S$ are found in the narrow interval $[\delta, 1]$ and the condition number of $\widetilde{S}_0^{-1}S$ is bounded by $\delta^{-1} \simeq 1.1$, i.e. very close to unity. This result is an improvement of the results in [1], [2].

The preconditioner requires two solutions with matrix $M + \gamma A$ at each iteration. Besides this and some matrix vector products, to evaluate the action of the Schur complement, this method requires also an inner solution with matrix M.

3. A transformed matrix approach

3.1. The transformed preconditioner. We consider now an alternative to the Schur complement method where only the two systems with linear combinations of M and A are needed per iteration, i.e. the additional solution with a system matrix M is not required. The next lemma will be useful.

Lemma 3.1. *Let $ab > 0$ and assume that $A + \sqrt{ab}B_i$, $i = 1, 2$, are invertible. Then a matrix in the form*

$$\mathcal{B} = \begin{bmatrix} A + \sqrt{ab}(B_1 + B_2) & -aB_2 \\ bB_1 & A \end{bmatrix} \tag{3.1}$$

can be written in the transformed form

$$\mathcal{B} = \begin{bmatrix} I & 0 \\ \alpha I & I \end{bmatrix} \begin{bmatrix} A + \sqrt{ab}B_1 & -aB_2 \\ 0 & A + \sqrt{ab}B_2 \end{bmatrix} \begin{bmatrix} I & 0 \\ -\alpha I & I \end{bmatrix}, \tag{3.2}$$

where $\alpha = \sqrt{b/a}$, *which shows that* $\mathcal{B}$ *is nonsingular.*

P r o o f. Since $a\alpha = \sqrt{ab}$, the above matrix equals

$$\begin{bmatrix} I & 0 \\ \alpha I & I \end{bmatrix} \begin{bmatrix} A + \sqrt{ab}B_1 + a\alpha B_2 & -aB_2 \\ -\alpha(A + \sqrt{ab}B_2) & A + \sqrt{ab}B_2 \end{bmatrix}$$

$$= \begin{bmatrix} A + \sqrt{ab}(B_1 + B_2) & -aB_2 \\ \alpha\sqrt{ab}B_1 & A + (\sqrt{ab} - \alpha a)B_2 \end{bmatrix}$$

$$= \begin{bmatrix} A + \sqrt{ab}(B_1 + B_2) & -aB_2 \\ bB_1 & A \end{bmatrix} = \mathcal{B}.$$

□

It follows that a system

$$\mathcal{B} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

can be solved as

$$\begin{bmatrix} A + \sqrt{ab}B_1 & -aB_2 \\ 0 & A + \sqrt{ab}B_2 \end{bmatrix} \begin{bmatrix} x_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} f_1 \\ g_2 \end{bmatrix}, \tag{3.3}$$

where $y_2 = x_2 - \sqrt{b/a}x_1$, $g_2 = f_2 - \sqrt{b/a}f_1$. Hence, besides some vector additions, it requires only one solution with each of $H_i = A + \sqrt{ab}B_i$, $i = 1, 2$, namely with $H_2 = A + \sqrt{ab}B_2$ to compute y_2, followed by a matrix vector multiplication with aB_2 and a solution with $H_1 = A + \sqrt{ab}B_1$.

Depending on the form of the given system with matrix $\mathcal{A}$, we modify typically the pivot block to get a matrix on the same form as $\mathcal{B}$ in (3.1) and use it as a preconditioner when solving the given system with $\mathcal{A}$. As an example take first

$$\mathcal{A} = \begin{bmatrix} A & -aB_2 \\ bB_1 & A \end{bmatrix}.$$

Here we add $\sqrt{ab}(B_1 + B_2)$ to the pivot block to form the matrix $\mathcal{B}$.

Proposition 3.1. *Let* A *be symmetric and positive definite, let* $B_1 = B$, $B_2 = B^{\mathrm{T}}$, *and assume that* $B + B^{\mathrm{T}}$ *is positive semidefinite. Then the eigenvalues of the preconditioned matrix* $\mathcal{B}^{-1}\mathcal{A}$ *are contained in the interval* $[\frac{1}{2}, 1]$.

P r o o f. To analyse the eigenvalues (λ) of the preconditioned matrix $\mathcal{B}^{-1}\mathcal{A}$, we consider the generalized eigenvalue problem,

$$\lambda\mathcal{B}\begin{bmatrix} x \\ y \end{bmatrix} = \mathcal{A}\begin{bmatrix} x \\ y \end{bmatrix}, \quad \|x\| + \|y\| \neq 0$$

or

$$\mu\mathcal{A}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \sqrt{ab}(B + B^{\mathrm{T}}) & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix},$$

where $\mu = 1/\lambda - 1$.

It follows that $\mu = 0$ ($\lambda = 1$) if $x = 0$, $y \neq 0$. For $x \neq 0$, we have

$$Ay = -bBx$$

and

$$\mu(A + abB^{\mathrm{T}}A^{-1}B)x = \sqrt{ab}(B + B^{\mathrm{T}})x. \tag{3.4}$$

Since $B + B^{\mathrm{T}}$ is positive semidefinite, it follows that $\mu \geqslant 0$, i.e. $\lambda \leqslant 1$.

By use of the congruence transformation, $\widetilde{B} = A^{-1/2}BA^{-1/2}$, (3.4) can be written in the form

$$\mu(I + ab\widetilde{B}^{\mathrm{T}}\widetilde{B})\tilde{x} = \sqrt{ab}(\widetilde{B} + \widetilde{B}^{\mathrm{T}})\tilde{x},$$

where $\tilde{x} = A^{1/2}x$.

Since $(I - \sqrt{ab}\widetilde{B}^{\mathrm{T}})(I - \sqrt{ab}\widetilde{B}) \geqslant 0$, it follows that

$$I + ab\widetilde{B}^{\mathrm{T}}\widetilde{B} \geqslant \sqrt{ab}(\widetilde{B} + \widetilde{B}^{\mathrm{T}}).$$

Hence $\mu \leqslant 1$, so $\frac{1}{2} \leqslant \lambda \leqslant 1$. □

We show now that the above transformation in (3.2) allows also an efficient implementation of the preconditioned iterations. Besides solving systems by matrix $A + \sqrt{ab}B_i$, $i = 1, 2$, it involves namely even fewer matrix-vector multiplications than a matrix-vector multiplication by $\mathcal{A}$. To see this, we use

$$\mathcal{B}^{-1}\mathcal{A} = \mathcal{B}^{-1}(\mathcal{B} - (\mathcal{B} - \mathcal{A})) = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} - \sqrt{ab}\mathcal{B}^{-1}\begin{bmatrix} B_1 + B_2 & 0 \\ 0 & 0 \end{bmatrix}.$$

Here, by (3.2),

$$\begin{aligned}
&\mathcal{B}^{-1}\begin{bmatrix} B_1+B_2 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \\
&= \begin{bmatrix} I & 0 \\ \alpha I & I \end{bmatrix}\begin{bmatrix} (A+\sqrt{ab}B_1)^{-1} & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} A+\sqrt{ab}B_2 & aB_2 \\ 0 & I \end{bmatrix} \\
&\quad \begin{bmatrix} (A+\sqrt{ab}B_2)^{-1} & 0 \\ 0 & (A+\sqrt{ab}B_2)^{-1} \end{bmatrix}\begin{bmatrix} I & 0 \\ -\alpha I & I \end{bmatrix}\begin{bmatrix} (B_1+B_2)x \\ 0 \end{bmatrix} \\
&= \begin{bmatrix} I & 0 \\ \alpha I & I \end{bmatrix}\begin{bmatrix} (A+\sqrt{ab}B_1)^{-1} & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} A+\sqrt{ab}B_2 & aB_2 \\ 0 & I \end{bmatrix}\begin{bmatrix} \tilde{x} \\ -\alpha\tilde{x} \end{bmatrix},
\end{aligned}$$

where $\tilde{x} = (A+\sqrt{ab}B_2)^{-1}(B_1+B_2)x$. Hence,

$$\begin{aligned}
&\mathcal{B}^{-1}\begin{bmatrix} (B_1+B_2)x \\ 0 \end{bmatrix} = \begin{bmatrix} I & 0 \\ \alpha I & I \end{bmatrix}\begin{bmatrix} (A+\sqrt{ab}B_1)^{-1} & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} A\tilde{x} \\ -\alpha\tilde{x} \end{bmatrix} \\
&= \begin{bmatrix} I & 0 \\ \alpha I & I \end{bmatrix}\begin{bmatrix} (A+\sqrt{ab}B_1)^{-1}A\tilde{x} \\ -\alpha\tilde{x} \end{bmatrix} = \begin{bmatrix} I & 0 \\ \alpha I & I \end{bmatrix}\begin{bmatrix} \tilde{x}-\sqrt{ab}(A+\sqrt{ab}B_1)^{-1}B_1\tilde{x} \\ -\alpha\tilde{x} \end{bmatrix} \\
&= \begin{bmatrix} \tilde{x}-\sqrt{ab}(A+\sqrt{ab}B_1)^{-1}B_1\tilde{x} \\ -b(A+\sqrt{ab}B_1)^{-1}B_1\tilde{x} \end{bmatrix}.
\end{aligned}$$

Consider the solution of a system

$$\mathcal{A}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}.$$

It follows from the above that the computation of a preconditioned residual can be done as

$$\begin{bmatrix} r_1 \\ r_2 \end{bmatrix} = \mathcal{B}^{-1}\left(\mathcal{A}\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}\right) = \begin{bmatrix} x \\ y \end{bmatrix} - \mathcal{B}^{-1}\begin{bmatrix} f_1 \\ f_2 \end{bmatrix} + \begin{bmatrix} \sqrt{ab}(\tilde{z}-\tilde{x}) \\ b\tilde{z} \end{bmatrix},$$

where $\tilde{x} = (A+\sqrt{ab}B_2)^{-1}(B_1+B_2)x$, $\tilde{z} = \sqrt{ab}(A+\sqrt{ab}B_1)^{-1}B_1\tilde{x}$. Hence, besides a solution of a system with $(A+\sqrt{ab}B_2)$ followed by one with $(A+\sqrt{ab}B_1)$, the computation of a residual in each iteration, involves only a matrix multiplication by B_1 and one by B_1+B_2. In many applications, $B_2 = B_1^{\mathrm{T}}$, in which case B_1+B_2 is symmetric so a matrix-vector multiplication by it can be further simplified.

The initial computation of $\mathcal{B}^{-1}\begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$ takes place as in (3.3). Note also that the second component y of the vector does not enter in the evaluation of the last term in

the expression for the residual, which therefore does not need to be accessed during the further computation of (r_1, r_2).

We consider now some applications of the transformed matrix approach.

3.2. Optimal control problem. Consider the solution of the optimal control problem in Section 2.2, where we used a Schur complement reduction approach. We use now instead the transformed matrix approach to handle the solution of the arising saddle point type matrix,

$$\begin{bmatrix} M & K^{\mathrm{T}} \\ K & -\alpha M \end{bmatrix}.$$

We transform it first to

$$\begin{bmatrix} I & 0 \\ 0 & \alpha^{-1/2} I \end{bmatrix} \begin{bmatrix} M & K^{\mathrm{T}} \\ K & -\alpha M \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \alpha^{-1/2} I \end{bmatrix} = \begin{bmatrix} M & \alpha^{-1/2} K^{\mathrm{T}} \\ \alpha^{-1/2} K & -M \end{bmatrix}$$

to obtain the matrix $\mathcal{A} = \left[\begin{smallmatrix} M & \alpha^{-1/2}K^{\mathrm{T}} \\ \alpha^{-1/2}K & -M \end{smallmatrix}\right]$. This matrix is preconditioned by perturbing the pivot matrix block, to form

$$\mathcal{B} = \begin{bmatrix} M + \alpha^{-1/2}(K + K^{\mathrm{T}}) & \alpha^{-1/2} K^{\mathrm{T}} \\ \alpha^{-1/2} K & -M \end{bmatrix}.$$

As follows from Proposition 3.1, the resulting spectral condition number is bounded by 2, uniformly in the parameter α, $\alpha > 0$. The matrix $\mathcal{B}$ can be transformed to block triangular form by

$$\begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \begin{bmatrix} M + \alpha^{-1/2}(K + K^{\mathrm{T}}) & \alpha^{-1/2} K^{\mathrm{T}} \\ \alpha^{-1/2} K & -M \end{bmatrix} \begin{bmatrix} I & 0 \\ -I & -I \end{bmatrix} \tag{3.5}$$

$$\begin{bmatrix} I & 0 \\ -I & I \end{bmatrix} \begin{bmatrix} M + \alpha^{-1/2} K & -\alpha^{-1/2} K^{\mathrm{T}} \\ M + \alpha^{-1/2} K & M \end{bmatrix} = \begin{bmatrix} M + \alpha^{-1/2} K & -\alpha^{-1/2} K^{\mathrm{T}} \\ 0 & M + \alpha^{-1/2} K^{\mathrm{T}} \end{bmatrix}.$$

3.3. Radau time integration. As another application, consider next the Radau two-point integration matrix. Here

$$\mathcal{A} = \begin{bmatrix} M + \frac{5}{12}A & -\frac{1}{12}A \\ \frac{9}{12}A & M + \frac{3}{12}A \end{bmatrix} = \begin{bmatrix} D_1 & -\frac{1}{12}A \\ \frac{9}{12}A & D_2 \end{bmatrix}. \tag{3.6}$$

To form a preconditioner to $\mathcal{A}$ we first approximate $\mathcal{A}$ by $\hat{A} = \left[\begin{smallmatrix} M+\frac{3}{12}A & -\frac{1}{12}A \\ \frac{9}{12}A & M+\frac{3}{12}A \end{smallmatrix}\right]$ for which we construct the preconditioner

$$\mathcal{B} = \begin{bmatrix} M + \frac{3}{12}A + \frac{2\sqrt{9}}{12}A & -\frac{1}{12}A \\ \frac{9}{12}A & M + \frac{3}{12}A \end{bmatrix} = \begin{bmatrix} M + \frac{3}{4}A & -\frac{1}{12}A \\ \frac{9}{12}A & M + \frac{3}{12}A \end{bmatrix},$$

i.e. using the same approach as before. It follows that

$$\mathcal{B} = \mathcal{A} + \begin{bmatrix} \frac{1}{3}A & 0 \\ 0 & 0 \end{bmatrix}. \tag{3.7}$$

Proposition 3.2. *For the Radau time-integration method the eigenvalues of the preconditioned matrix $\mathcal{B}^{-1}\mathcal{A}$ are contained in the interval $[\frac{2}{3}, 1]$.*

P r o o f. Let $\mu = \frac{1}{\lambda} - 1$. Then

$$\mu\mathcal{A}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{1}{3}A & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}, \quad \|x\| + \|y\| \neq 0.$$

For $x \neq 0$, it follows

$$(M + \tfrac{3}{12}A)y = -\tfrac{9}{12}Ax$$

and

$$\mu[M + \tfrac{5}{12}A + \tfrac{9}{144}A(M + \tfrac{3}{12}A)^{-1}A]x = \tfrac{1}{3}Ax.$$

Using the congruence transformation, $\widetilde{A} = M^{-1/2}AM^{-1/2}$, we get now

$$\mu[I + \tfrac{5}{12}\widetilde{A} + \tfrac{1}{16}\widetilde{A}(I + \tfrac{3}{12}\widetilde{A})^{-1}\widetilde{A}]\hat{x} = \tfrac{1}{3}\widetilde{A}\hat{x},$$

where $\hat{x} = (I + \frac{1}{4}\widetilde{A})^{-1}x$, so

$$\mu[I + \tfrac{8}{12}\widetilde{A} + (\tfrac{1}{16} + \tfrac{15}{144})\widetilde{A}^2]\hat{x} = \tfrac{1}{3}\widetilde{A}(I + \tfrac{1}{4}\widetilde{A})\hat{x}$$

or

$$\mu(I + \tfrac{2}{3}\widetilde{A} + \tfrac{1}{6}\widetilde{A}^2)\hat{x} = \tfrac{1}{3}\widetilde{A}(I + \tfrac{1}{4}\widetilde{A})\hat{x},$$

i.e.

$$\mu[I + \tfrac{2}{3}\widetilde{A}(I + \tfrac{1}{4}\widetilde{A})]\hat{x} = \tfrac{1}{3}\widetilde{A}(I + \tfrac{1}{4}\widetilde{A})\hat{x}.$$

Hence, $0 \leqslant \mu \leqslant \frac{1}{2}$, $\frac{2}{3} \leqslant \lambda \leqslant 1$ and the condition number is bounded by $\frac{3}{2}$. □

This method requires the solution of two systems with matrix

$$M + \tfrac{3}{12}A + \sqrt{ab}2A = M + \tfrac{3}{4}A$$

per iteration. The method in Section 2, using preconditioning of the reduced Schur complement matrix requires one solution with matrix M and two solutions with matrices $M + \gamma B = M + \frac{1}{\sqrt{6}}A$, per iteration step (here A contains the factor τ (time step)). As we have shown in Section 2, it has an even smaller condition number.

This method requires however that M is positive definite. For Stokes-like problems, this does not hold, as the corresponding matrix is indefinite. In Section 4 we show that the transformation method is applicable also for such problems.

3.4. Complex valued systems solved in real arithmetics. Complex valued systems can be solved in real arithmetics, which saves demand of memory and computational complexity. We show that the transformed matrix approach gives a simpler and shorter derivation of the method and its properties than the approach taken in [6], see also [5] and [8].

Consider then the system

$$(A + \mathrm{i}B)(x + \mathrm{i}y) = f + \mathrm{i}g, \tag{3.8}$$

where A, B, x, y, f and g are real valued. We rewrite (3.8) in the form

$$\mathcal{A}\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f \\ g \end{bmatrix},$$

where $\mathcal{A} = \left[\begin{smallmatrix} A & -B \\ B & A \end{smallmatrix}\right]$ and, following the transformed matrix approach, precondition $\mathcal{A}$ by $\mathcal{B} = \left[\begin{smallmatrix} A+2B & -B \\ B & A \end{smallmatrix}\right]$. Let $\lambda(C)$ denote the eigenvalues of a matrix C.

Proposition 3.3. *Assume that $A + B$ is nonsingular. Then $\mathcal{B}$ is nonsingular, $\lambda(\mathcal{B}^{-1}\mathcal{A}) = 1$ for eigenvector $(\hat{x}, \hat{y})$, $\hat{x} \in \mathcal{N}(B)$ and the remaining eigenvalues satisfy*

$$\lambda(\mathcal{B}^{-1}\mathcal{A}) = 1 + 2[((A + B)^{-1}B)^2 - (A + B)^{-1}B].$$

P r o o f. The generalized eigenvalue problem

$$\lambda\mathcal{B}\begin{bmatrix} x \\ y \end{bmatrix} = \mathcal{A}\begin{bmatrix} x \\ y \end{bmatrix}, \quad \|x\| + \|y\| \neq 0,$$

can be rewritten in the form

$$\lambda\begin{bmatrix} I & 0 \\ -I & I \end{bmatrix}\mathcal{B}\begin{bmatrix} I & 0 \\ I & I \end{bmatrix}\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix}\mathcal{A}\begin{bmatrix} I & 0 \\ I & I \end{bmatrix}\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix},$$

where

$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = \begin{bmatrix} I & 0 \\ -I & I \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}.$$

Based on this relation a computation shows that

$$\lambda\begin{bmatrix} A + B & -B \\ 0 & A + B \end{bmatrix}\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = \begin{bmatrix} A - B & -B \\ 2B & A + B \end{bmatrix}\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} \tag{3.9}$$

or

$$(\lambda-1)\begin{bmatrix} A+B & -B \\ 0 & A+B \end{bmatrix}\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = 2\begin{bmatrix} -B & 0 \\ B & 0 \end{bmatrix}\begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix}.$$

Hence $\lambda = 1$ for eigenvectors $\tilde{x} \in \mathcal{N}(B)$, the nullspace of B. Let $\hat{y} = (\lambda-1)y$. For $\lambda \neq 1$ we have $\hat{y} = 2(A+B)^{-1}B\tilde{x}$ and

$$(\lambda-1)(A+B)\tilde{x} = 2(B(A+B)^{-1}B - B)\tilde{x},$$

which shows the result. □

Corollary 3.1. *Assume that A and B are symmetric and positive semidefinite and that $\mathcal{N}(A) \cap \mathcal{N}(B) = \{0\}$. Then the eigenvalues of $\mathcal{B}^{-1}\mathcal{A}$ are located in the interval $[\frac{1}{2}, 1]$.*

P r o o f. Let $\hat{x} = (A+B)^{1/2}\tilde{x}$, $\hat{B} = (A+B)^{-1/2}B(A+B)^{-1/2}$. Then it follows from Proposition 3.3 that

$$(\lambda-1)\hat{x}^{\mathrm{T}}\hat{x} = 2\hat{x}^{\mathrm{T}}(\hat{B}^2 - \hat{B})\hat{x}.$$

Clearly $0 \leqslant \hat{B} \leqslant I$. Since $0 \leqslant x - x^2 \leqslant \frac{1}{4}$ for $0 \leqslant x < 1$, it follows that $-\frac{1}{2} \leqslant \lambda - 1 \leqslant 0$. □

4. The transformed matrix approach for a time-dependent Stokes problem

We present first the problem and show how the discrete system can be reduced to a set of two equations. We discuss then the preconditioning method and a method based on a projection matrix.

4.1. Problem formulation and reduced systems. For a bounded domain Ω the time dependent Stokes equation is given by

$$\begin{cases} \dfrac{\partial \underline{u}}{\partial t} - \Delta \underline{u} + \nabla p = f, \\ \nabla \cdot \underline{u} = 0 \qquad\qquad\qquad \text{in } \Omega \times [0, \tau], \end{cases}$$

and $\underline{u} = \underline{u}_D$ on $\partial\Omega$, and a given initial value $\underline{u}(0) = \underline{u}_0$.

Here $\underline{u}$ is the velocity field and p is the pressure. After a proper inf-sup, stable space discretization, and use of the two-point Radau time-integration method, the

resulting system for the first time step (and similarly for the following time steps) takes the form (for notational simplicity we let here $\tau = 1$)

$$
\left\{
\begin{aligned}
&M\underline{u}(\tfrac{1}{3}) + \tfrac{5}{12}(K\underline{u}(\tfrac{1}{3}) + B^{\mathrm{T}}p(\tfrac{1}{3})) - \tfrac{1}{12}[K\underline{u}(1) + B^{\mathrm{T}}p(1)] \\
&\qquad = M\underline{u}_0 + \tfrac{\tau}{12}(5f_1 - f_2), \\
&\tfrac{5}{12}B\underline{u}(\tfrac{1}{3}) - \tfrac{1}{12}B\underline{u}(1) = 0, \\
&M\underline{u}(1) + \tfrac{9}{12}\left(K\underline{u}(\tfrac{1}{3}) + B^{\mathrm{T}}p(\tfrac{1}{3})\right) + \tfrac{3}{12}(K\underline{u}(1) + B^{\mathrm{T}}p(1)) \\
&\qquad = M\underline{u}_0 + \tfrac{\tau}{4}(3f_1 + f_2), \\
&\tfrac{9}{12}B\underline{u}(\tfrac{1}{3}) + \tfrac{3}{12}B\underline{u}(1) = 0.
\end{aligned}
\right. \tag{4.1}
$$

Here M is the block mass-matrix, $K = \tau\widetilde{K}$, where $\widetilde{K}$ is the discrete negative diffusion matrix, $B = \tau\widetilde{B}$ and $B^{\mathrm{T}} = \tau\widetilde{B}^{\mathrm{T}}$, where $\widetilde{B}$, $\widetilde{B}^{\mathrm{T}}$ are the discrete divergence and gradient operators, respectively and $\tau > 0$ is the time step. We assume that the so called inf-sup stable finite element pairs (see e.g. [11]) have been used, which implies that the matrix B has full rank. The part of the matrix system corresponding to the divergence constraint can be written as

$$
\begin{bmatrix} \frac{5}{12} & -\frac{1}{12} \\ \frac{9}{12} & \frac{3}{12} \end{bmatrix}
\begin{bmatrix} B\underline{u}(\frac{1}{3}) \\ B\underline{u}(1) \end{bmatrix}
= \begin{bmatrix} 0 \\ 0 \end{bmatrix}
$$

and this matrix is nonsingular. Hence, it follows that $B\underline{u}(\frac{1}{3}) = 0$, $B\underline{u}(1) = 0$, that is, the discrete divergence constraint holds for the solution in both integration points. We show now that this enables one to reduce the system to a two-by-two block matrix form.

Let $\underline{x}_1 = \underline{u}(\frac{1}{3})$, $\underline{x}_2 = \underline{u}(1)$, $y_1 = p(\frac{1}{3})$, and $y_2 = p(1)$. Then the remaining equations in (4.1) can be written as

$$
\left\{
\begin{aligned}
&M\underline{x}_1 + \tfrac{5}{12}(K\underline{x}_1 + B^{\mathrm{T}}y_1) - \tfrac{1}{12}(K\underline{x}_2 + B^{\mathrm{T}}y_2) = g_1, \\
&\tfrac{9}{12}(K\underline{x}_1 + B^{\mathrm{T}}y_1) + M\underline{x}_2 + \tfrac{3}{12}(K\underline{x}_2 + B^{\mathrm{T}}y_2) = g_2,
\end{aligned}
\right. \tag{4.2}
$$

where g_1, g_2 are the corresponding r.h.s. expressions in (4.1). By multiplication of each row in (4.2) by BM^{-1}, and using $\underline{x}_i \in \mathcal{N}(B)$, we get

$$
\left\{
\begin{aligned}
&\tfrac{5}{12}(BM^{-1}K\underline{x}_1 + BM^{-1}B^{\mathrm{T}}y_1) - \tfrac{1}{12}(BM^{-1}K\underline{x}_2 + BM^{-1}B^{\mathrm{T}}y_2) = \tilde{g}_1, \\
&\tfrac{9}{12}(BM^{-1}K\underline{x}_1 + BM^{-1}B^{\mathrm{T}}y_1) + \tfrac{3}{12}(BM^{-1}K\underline{x}_2 + BM^{-1}B^{\mathrm{T}}y_2) = \tilde{g}_2,
\end{aligned}
\right.
$$

where $\tilde{g}_i = BM^{-1}g_i$, $i = 1, 2$.

Here we eliminate first the pair (x_2, y_2) and then the pair (x_1, y_1) to get

$$
\left\{
\begin{aligned}
&BM^{-1}K\underline{x}_1 + BM^{-1}B^{\mathrm{T}}y_1 = \hat{g}_1 := \tfrac{3}{2}(\tilde{g}_1 + \tfrac{1}{3}\tilde{g}_2), \\
&BM^{-1}K\underline{x}_2 + BM^{-1}B^{\mathrm{T}}y_2 = \hat{g}_2 := \tfrac{1}{2}(5\tilde{g}_2 - 9\tilde{g}_1).
\end{aligned}
\right.
$$

Since B has full rank, it follows that $BM^{-1}B^{\mathrm{T}}$ is nonsingular so that

$$y_i = -(BM^{-1}B^{\mathrm{T}})^{-1}[BM^{-1}K\underline{x}_i - \hat{g}_i], \quad i = 1, 2. \tag{4.3}$$

Substituting this into (4.2), we get

$$\begin{cases} M\underline{x}_1 + \frac{5}{12}(I-P)K\underline{x}_1 - \frac{1}{12}(I-P)K\underline{x}_2 = h_1 \\ \frac{9}{12}(I-P)K\underline{x}_1 + M\underline{x}_2 + \frac{3}{12}(I-P)K\underline{x}_2 = h_2 \end{cases} \tag{4.4}$$

for some right hand sides h_i, $i = 1, 2$. (Actually $h_i = (I-P)g_i$, $i = 1, 2$.) Here

$$P = B^{\mathrm{T}}(BM^{-1}B^{\mathrm{T}})^{-1}BM^{-1}$$

is a projection matrix onto the range of B^{T}. Here we can apply a Richardson iteration method, multiplying both equations by M^{-1}. It follows that iterations with this matrix will only occur for vectors $\underline{x}$ in the orthogonal complement subspace of range(B^{T}). For Stokes problem it means that convergence will take place in the subspace of divergence-free vectors. Besides several solutions with the mass matrix, the method involves computing actions of P. For this purpose, we can replace M in $BM^{-1}B^{\mathrm{T}}$ by its diagonal part and form the so arising version of $BM^{-1}B^{\mathrm{T}}$ explicitly and then use some standard available preconditioned iterative solution method to compute actions of $(BM^{-1}B^{\mathrm{T}})^{-1}$. However, since this can be costly, in the next subsection we present an alternative approach to solve time-dependent Stokes problems.

Note that a multiplication of the equations in (4.4) by BM^{-1} gives $B\underline{x}_i = 0$, $i = 1, 2$, which is in accordance with the validity of the divergence-free condition.

4.2. Block matrix preconditioner based on the transformed matrix approach. Since computing actions of matrix P can be costly, it may not be feasible to use the previous equation to solve for $\underline{x}_1$, $\underline{x}_2$. However, we can compare it with (3.6) to see how a preconditioning similar to (3.7) will function. Write then the global system matrix (4.1) in the form

$$\mathcal{A} = \begin{bmatrix} \hat{M} + \frac{5}{12}\hat{K} & -\frac{1}{12}\hat{K} \\ \frac{9}{12}\hat{K} & \hat{M} + \frac{3}{12}\hat{K} \end{bmatrix},$$

where $\hat{M} = \begin{bmatrix} M & 0 \\ 0 & 0 \end{bmatrix}$, $\hat{K} = \begin{bmatrix} K & B^{\mathrm{T}} \\ B & 0 \end{bmatrix}$, and where $\hat{M}$ is positive semidefinite.

As shown in (3.7), the preconditioner

$$\mathcal{B} = \begin{bmatrix} \hat{M} + \frac{9}{12}\hat{K} & -\frac{1}{12}\hat{K} \\ \frac{9}{12}\hat{K} & \hat{M} + \frac{3}{12}\hat{K} \end{bmatrix}$$

gives eigenvalue bounds of $\mathcal{B}^{-1}\mathcal{A}$ in the interval $\left[\frac{2}{3}, 1\right]$. Hence, there will be very few outer iterations. Further, it was shown that $\mathcal{B}$ can be written in the transformed form

$$\mathcal{B} = \begin{bmatrix} I & 0 \\ \alpha & I \end{bmatrix} \begin{bmatrix} \hat{M} + \frac{6}{12}\hat{K} & -\frac{1}{12}\hat{K} \\ 0 & \hat{M} + \frac{6}{12}\hat{K} \end{bmatrix} \begin{bmatrix} I & 0 \\ -\alpha & I \end{bmatrix},$$

where $\alpha = \sqrt{ab} = \sqrt{9/1} = 3$. This enables us to solve systems with matrix $\mathcal{B}$ that involves just two solutions with matrix $\hat{M} + \frac{1}{2}\hat{K}$. This is the matrix that appears in the inner iteration method.

They can be solved as a time-dependent Stokes problem with matrix

$$\begin{bmatrix} M + \frac{1}{2}K & \frac{1}{2}B^{\mathrm{T}} \\ \frac{1}{2}B & 0 \end{bmatrix}$$

for the different right-hand sides that appear in the outer iteration method. For this purpose one can use an inner iteration method with the preconditioner

$$\begin{bmatrix} M + \frac{1}{2}K & 0 \\ \frac{1}{2}B & -\widetilde{S} \end{bmatrix}, \tag{4.5}$$

where $\widetilde{S}$ is an approximation of the Schur complement matrix

$$S = B(M + \tfrac{1}{2}K)^{-1}B^{\mathrm{T}}.$$

A straightforward computation shows that for a matrix $\widetilde{B}$ with full row rank,

$$(\widetilde{B}(I + \widetilde{B}^{\mathrm{T}}\widetilde{B})^{-1}\widetilde{B}^{\mathrm{T}})^{-1} = I + (\widetilde{B}\widetilde{B}^{\mathrm{T}})^{-1}. \tag{4.6}$$

Hence, letting $\widetilde{B} = W^{-1/2}BM^{-1/2}$, we obtain

$$(B(M + B^{\mathrm{T}}W^{-1}B)^{-1}B^{\mathrm{T}})^{-1} = W^{-1} + (BM^{-1}B^{\mathrm{T}})^{-1},$$

where W is an spd matrix.

Hence, if matrix $K = B^{\mathrm{T}}M_p^{-1}B$, where M_p is the pressure mass matrix, or if K can be approximated by a matrix in such a form, then with $W = 2M_p$, it follows from (4.6) that

$$S^{-1} = (B(M + \tfrac{1}{2}B^{\mathrm{T}}M_p^{-1}B)^{-1}B^{\mathrm{T}})^{-1} = \tfrac{1}{2}M_p^{-1} + (BM^{-1}B^{\mathrm{T}})^{-1}.$$

Here $BM^{-1}B^{\mathrm{T}}$ can be approximated by the negative scalar Laplacian. Therefore,

$$\widetilde{S}^{-1} = \frac{1}{2}M_p^{-1} + (-\Delta)^{-1} \tag{4.7}$$

defines an efficient approximation of the inverse of the Schur complement matrix S, to be used in the preconditioned matrix (4.5), see also [13] and [2]. Normally the action of both of the inverses can be computed efficiently.

Note that K, B, and B^{T} contain the factor τ. Hence, the second term in (4.7) contains a factor τ^{-1} larger than the first term.

5. The PMHSS iteration method

In [9] a preconditioned iteration method, named the preconditioned modified Hermitian sequential subspace (PMHSS) method, is presented which also deals with solution of systems of the form we consider and with a preconditioning matrix that involves solving systems with a matrix in the form of a linear combination of the block row matrices in the given matrix. We show here a much simplified and more general analysis of the corresponding spectrum and rate of convergence of that method. As in our method, the presentation in [9] involves also a method which involves a preconditioner that contains a linear combination of one of the block row matrices with possibly another matrix that is not necessarily equal to the second row block matrix. However, the analysis in [9] indicates that it cannot considerably improve the preconditioning and we let it here be the other block row matrix only.

Let $\alpha > 0$ be a parameter and let $\mathcal{A} = \left[\begin{smallmatrix} A & -B \\ B & A \end{smallmatrix}\right]$ and $\mathcal{A}\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right] = \left[\begin{smallmatrix} f_1 \\ f_2 \end{smallmatrix}\right]$ be the given system to be solved, where A and B are symmetric and positive semidefinite and $\mathcal{N}(A) \cap \mathcal{N}(B) = \{0\}$. The method can be written as an alternating fixed-point iteration method, in the form

$$\begin{bmatrix} (\alpha+1)I & 0 \\ 0 & (\alpha+1)I \end{bmatrix} \begin{bmatrix} Ay^{(k+1/2)} \\ Az^{(k+1/2)} \end{bmatrix} = \begin{bmatrix} \alpha A & B \\ -B & \alpha A \end{bmatrix} \begin{bmatrix} y^{(k)} \\ z^{(k)} \end{bmatrix} + \begin{bmatrix} f_1 \\ f_2 \end{bmatrix},$$

$$\begin{bmatrix} \alpha A + B & 0 \\ 0 & \alpha A + B \end{bmatrix} \begin{bmatrix} y^{(k+1)} \\ z^{(k+1)} \end{bmatrix} = \begin{bmatrix} \alpha A & -A \\ A & \alpha A \end{bmatrix} \begin{bmatrix} y^{(k+1/2)} \\ z^{(k+1/2)} \end{bmatrix} + \begin{bmatrix} f_2 \\ -f_1 \end{bmatrix}, \quad k = 0, 1, \ldots$$

Let $\underline{x}^{(k)} = \left[\begin{smallmatrix} y^{(k)} \\ z^{(k)} \end{smallmatrix}\right]$. Since

$$\begin{bmatrix} (\alpha+1)^{-1}I & 0 \\ 0 & (\alpha+1)^{-1}I \end{bmatrix} \begin{bmatrix} \alpha I & -I \\ I & \alpha I \end{bmatrix} = \frac{1}{\alpha+1} \begin{bmatrix} \alpha I & -I \\ I & \alpha I \end{bmatrix},$$

it follows that the method can be reformulated as a Richardson iteration method,

$$\underline{x}^{(k+1)} = L(A;\alpha)\underline{x}^{(k)} + R(A;\alpha)\underline{g}, \quad k = 0, 1, \ldots$$

for some block vector $\underline{g}$, where

$$L(A;\alpha) = \frac{1}{\alpha+1}\begin{bmatrix}(\alpha A+B)^{-1} & 0\\ 0 & (\alpha A+B)^{-1}\end{bmatrix}\begin{bmatrix}\alpha I & -I\\ I & \alpha I\end{bmatrix}\begin{bmatrix}\alpha A & B\\ -B & \alpha A\end{bmatrix} \tag{5.1}$$
$$= \frac{1}{\alpha+1}\begin{bmatrix}(\alpha A+B)^{-1} & 0\\ 0 & (\alpha A+B)^{-1}\end{bmatrix}\begin{bmatrix}\alpha^2 A+B & \alpha(B-A)\\ \alpha(A-B) & \alpha^2 A+B\end{bmatrix}$$

and

$$R(A;\alpha) = \frac{\alpha}{\alpha+1}\begin{bmatrix}(\alpha A+B)^{-1} & 0\\ 0 & (\alpha A+B)^{-1}\end{bmatrix}\begin{bmatrix}I & I\\ -I & I\end{bmatrix}.$$

Proposition 5.1. *Let $\alpha > 0$, $\alpha A + B$ be positive definite. Then the eigenvalues (λ) of $L(A;\alpha)$ are bounded as follows:*

$$\frac{1}{\alpha+1}\min(\alpha,1) \leqslant \mathrm{Re}(\lambda) \leqslant \frac{1}{\alpha+1}\max(\alpha,1)$$

and

$$|\mathrm{Im}(\lambda)| \leqslant \frac{1}{\alpha+1}\max(1,\alpha).$$

P r o o f. Let α be a fixed positive number. In order to compute bounds for the eigenvalues of the iteration matrix $L(A;\alpha)$ we consider the generalized eigenvalue problem

$$\lambda\begin{bmatrix}\alpha A+B & 0\\ 0 & \alpha A+B\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = \frac{1}{\alpha+1}\begin{bmatrix}\alpha^2 A+B & \alpha(B-A)\\ \alpha(A-B) & \alpha^2 A+B\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix}, \quad |x|+|y|=1,$$

or

$$(\alpha+1)\lambda\begin{bmatrix}E & 0\\ 0 & E\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix} = \begin{bmatrix}(\alpha^2-\alpha)A+E & \alpha E-(\alpha^2+\alpha)A\\ (\alpha^2+\alpha)A-\alpha E & (\alpha^2-\alpha)A+E\end{bmatrix}\begin{bmatrix}x\\ y\end{bmatrix},$$

where $E = \alpha A + B$. Since, by assumption, E is symmetric and positive definite, we can apply the congruence transformation,

$$(\alpha+1)\lambda\begin{bmatrix}\tilde{x}\\ \tilde{y}\end{bmatrix} = \begin{bmatrix}(\alpha-1)\widetilde{A}+I & \alpha I-(\alpha+1)\widetilde{A}\\ (\alpha+1)\widetilde{A}-\alpha I & (\alpha-1)\widetilde{A}+I\end{bmatrix}\begin{bmatrix}\tilde{x}\\ \tilde{y}\end{bmatrix},$$

where $\widetilde{A} = E^{-1/2}(\alpha A)E^{-1/2}$, $\tilde{x} = E^{1/2}x$, $\tilde{y} = E^{1/2}y$. Hence,

$$\Big(\lambda - \frac{1}{\alpha+1}\Big)\begin{bmatrix}\tilde{x}\\ \tilde{y}\end{bmatrix} = \begin{bmatrix}\frac{\alpha-1}{\alpha+1}\widetilde{A} & \frac{\alpha}{\alpha+1}I-\widetilde{A}\\ \widetilde{A}-\frac{\alpha}{\alpha+1}I & \frac{\alpha-1}{\alpha+1}\widetilde{A}\end{bmatrix}\begin{bmatrix}\tilde{x}\\ \tilde{y}\end{bmatrix}$$

or

$$\Big(\lambda - \frac{1}{\alpha+1}\Big)\begin{bmatrix}\tilde{x}\\ \tilde{y}\end{bmatrix} = \frac{\alpha-1}{\alpha+1}\begin{bmatrix}\widetilde{A} & 0\\ 0 & \widetilde{A}\end{bmatrix}\begin{bmatrix}\tilde{x}\\ \tilde{y}\end{bmatrix} + \begin{bmatrix}0 & \frac{\alpha}{\alpha+1}I-\widetilde{A}\\ \widetilde{A}-\frac{\alpha}{\alpha+1}I & 0\end{bmatrix}\begin{bmatrix}\tilde{x}\\ \tilde{y}\end{bmatrix}, \tag{5.2}$$

where the matrix has been split in a symmetric positive semidefinite and a skew symmetric term.

It follows from (5.2) for any fixed α that the eigenvalues λ depend on μ as

$$\mathrm{Re}(\lambda) = \frac{1}{\alpha+1} + \frac{\alpha-1}{\alpha+1}\mu, \quad |\mathrm{Im}(\lambda)| = \Big|\frac{\alpha}{\alpha+1} - \mu\Big|,$$

where μ is an eigenvalue of $\widetilde{A}$. Since A and B are symmetric and positive semidefinite, it follows from $\mu(\alpha A + B)z = \alpha A z$ that $0 < \mu \leqslant 1$. This implies the stated bounds. □

It is readily seen that $\alpha = 1$ is an optimal choice. In this case $\lambda = \frac{1}{2} \pm \mathrm{i}\,\mathrm{Im}(\lambda)$, where $|\mathrm{Im}(\lambda)| \leqslant \frac{1}{2}$ and $|\lambda| = \frac{1}{2}\sqrt{2}$. This agrees with the bound of the spectral radius found in [9].

As has been pointed out in [9], the corresponding eigenvector matrix in the PMHSS method is unitary in a certain inner product space, which for a normal two-by-two block matrix is unitary in the Euclidean space. This makes the method ideal for a MINRES iteration method.

As shown in [2], [10], see also references therein, for a related symmetric linear system the eigenvalues are real and contained in the interval $[-1, -\frac{1}{2}\sqrt{2}] \cup [\frac{1}{2}\sqrt{2}, 1]$. Hence, the method can be effective and is mesh independent. However, since there are both positive and negative eigenvalues, it will converge slower than the methods presented in the previous sections.

6. Iterative solution methods

For the solution of symmetric indefinite problems one can apply the MINRES method [16]. This requires, however, use of a symmetric positive definite preconditioner, typically a block diagonal matrix.

For optimal control problems this will be less efficient than the use of the preconditioner proposed in this paper which, however, is nonsymmetric so it requires the use of a GMRES or GCG-LS [7] method. Normally one solves the matrix systems arising in the preconditioner with inner iterations which implies a slightly variable preconditioner. Hence, the flexible version, FGMRES [17] or the variable preconditioned version of GCG-LS must be used anyway.

What is important then is to use a preconditioner that leads to a full eigenvector space, i.e. a preconditioned matrix that is normal. For completeness of the paper we recall now the reason for that.

The GMRES and GCG-LS methods result in iterative approximate vectors $x^k \in x^0 + \mathrm{span}\{r^0, Cr^0, \ldots, C^{k-1}r^0\}$, where x^0 is the initial approximation, r^0 is the corresponding residual and $C = \mathcal{B}^{-1}\mathcal{A}$ is the preconditioned matrix. This implies that the

residuals of the iteration vectors satisfy $r^k = \mathcal{P}_k(C)r^0$ for some polynomial $\mathcal{P}_k$ of degree k, which is normalized $\mathcal{P}_k(0) = 1$ at the origin. If C has a complete eigenvector space (or, at least if r^0 is represented by a subset of linearly independent eigenvectors of C), then the methods produce a best polynomial approximation on the set of eigenvalues $\{\lambda_i\}$ and $\min_{P_k} \max_{\lambda_i} |P_k(\lambda_i)|$ gives an upper bound of the convergence factor.

If this does not hold, i.e. if the eigenvector space is incomplete, then it can be seen that there exist initial residuals where this factor decays arbitrarily slow, see for instance [14]. Even if r^0 can be represented by this set of eigenvectors but the eigenvector space is incomplete then rounding errors, occurring during the iterations, introduce components outside this set and the rate of convergence of the method can slow down considerably.

We now show that the preconditioning matrix used in Section 3 has a complete eigenvector space so the above problematic issues cannot occur. Consider then the splitting

$$\mathcal{A} = \mathcal{B} - \sqrt{ab} \begin{bmatrix} B_1 + B_2 & 0 \\ 0 & 0 \end{bmatrix}.$$

We assume that $B_1 = B$, $B_2 = B^{\mathrm{T}}$. Then, using the explicit expression for $\mathcal{B}^{-1}$, we obtain

$$\mathcal{C} = \mathcal{B}^{-1}\mathcal{A} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix} - \begin{bmatrix} 0 & F \\ 0 & E \end{bmatrix}, \tag{6.1}$$

where $F = \sqrt{b/a}(I - H_2^{-1}A)H_1^{-1}(B_1 + B_2)$, and $E = \sqrt{ab}H_2^{-1}AH_1^{-1}(B_1 + B_2)$ and $H_1 = A + \sqrt{ab}B$, $H_2 = A + \sqrt{ab}B^{\mathrm{T}}$. Here E can be symmetrized by the similarity transformation

$$A^{-1/2}H_2EH_2^{-1}A^{1/2} = \sqrt{ab}A^{1/2}H_1^{-1}(B + B^{\mathrm{T}})H_1^{-1}A^{1/2}.$$

Further, since

$$H_1A^{-1}H_2 = (A + \sqrt{ab}B)A^{-1}(A + \sqrt{ab}B^{\mathrm{T}}) = A + \sqrt{ab}(B + B^{\mathrm{T}}) + abBA^{-1}B^{\mathrm{T}},$$

it follows by the Cauchy-Schwarz inequality that E has eigenvalues in the interval $[0, \frac{1}{2}]$. Hence, by (6.1), $\mathcal{B}^{-1}\mathcal{A} = \left[\begin{smallmatrix} I & -F \\ 0 & I-E \end{smallmatrix}\right]$ and has eigenvalues in the interval $[\frac{1}{2}, 1]$. To find its eigenvectors, consider the eigenvalue problem

$$\begin{bmatrix} I & -F \\ 0 & I - E \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}.$$

Here $\lambda = 1$ if $x \neq 0$, $y = 0$ or if $x = 0$, $Ey = 0$. For $\lambda \neq 1$ we have $Ey = (1-\lambda)y$ and $x = Fy/(1-\lambda)$. Since E has a complete eigenvector space, so has the preconditioned matrix C, i.e. C is a normal matrix.

7. Concluding remarks

Two types of methods to solve linear matrix systems in two-by-two block form with square matrix blocks have been analysed.

Although the Schur complement reduction method involves an additional solution with the pivot matrix block we have seen an example, the Radau time integration, where the matrix can be preconditioned efficiently with a resulting condition number very close to unity. The preconditioner involves just solving twice a system which is a linear combination of the mass matrix and the stiffness matrix.

The transformed method does not involve the additional mass matrix block and is applicable also for problems like time-dependent Stokes equation where an additional constraint (divergence-free flow), i.e. one leading to an indefinite submatrix, arises.

Both methods can be used for the solution of optimal control problems involving a partial differential equation as constraint. The condition number of the preconditioned matrix is bounded, typically by 2 or less, which holds uniformly with respect to the various problem parameters. The methods compete favourably with another type of approach, published in [9].

A c k n o w l e d g e m e n t . Comments by two referees on the original version of this paper helped to improve the presentation of it and is gratefully acknowledged.

References

[1] *O. Axelsson*: On the efficiency of a class of A-stable methods. BIT, Nord. Tidskr. Inf.-behandl. *14* (1974), 279–287. zbl MR doi

[2] *O. Axelsson, R. Blaheta, R. Kohut*: Preconditioning methods for high-order strongly stable time integration methods with an application for a DAE problem. Numer. Linear Algebra Appl. *22* (2015), 930–949. zbl MR doi

[3] *O. Axelsson, S. Farouq, M. Neytcheva*: Comparison of preconditioned Krylov subspace iteration methods for PDE-constrained optimization problems. Poisson and convection-diffusion control. Numer. Algorithms *73* (2016), 631–663. zbl MR doi

[4] *O. Axelsson, S. Farouq, M. Neytcheva*: Comparison of preconditioned Krylov subspace iteration methods for PDE-constrained optimization problems. Stokes control. Numer. Algorithms *74* (2017), 19–37. zbl MR doi

[5] *O. Axelsson, A. Kucherov*: Real valued iterative methods for solving complex symmetric linear systems. Numer. Linear Algebra Appl. *7* (2000), 197–218. zbl MR doi

[6] *O. Axelsson, M. Neytcheva, B. Ahmad*: A comparison of iterative methods to solve complex valued linear algebraic systems. Numer. Algorithms *66* (2014), 811–841. zbl MR doi

[7] *O. Axelsson, P. S. Vassilevski*: A black box generalized conjugate gradient solver with inner iterations and variable-step preconditioning. SIAM J. Matrix Anal. Appl. *12* (1991), 625–644. zbl MR doi
[8] *Z.-Z. Bai*: On preconditioned iteration methods for complex linear systems. J. Eng. Math. *93* (2015), 41–60. zbl MR doi
[9] *Z.-Z. Bai, M. Benzi, F. Chen, Z.-Q. Wang*: Preconditioned MHSS iteration methods for a class of block two-by-two linear systems with applications to distributed control problems. IMA J. Numer. Anal. *33* (2013), 343–369. zbl MR doi
[10] *Z.-Z. Bai, F. Chen, Z.-Q. Wang*: Additive block diagonal preconditioning for block two-by-two linear systems of skew-Hamiltonian coefficient matrices. Numer. Algorithms *62* (2013), 655–675. zbl MR doi
[11] *F. Brezzi, M. Fortin*: Mixed and Hybrid Finite Element Methods. Springer Series in Computational Mathematics 15, Springer, New York, 1991. zbl MR doi
[12] *J. C. Butcher*: Numerical Methods for Ordinary Differential Equations. John Wiley & Sons, Chichester, 2008. zbl MR doi
[13] *J. Cahouet, J.-P. Chabard*: Some fast 3D finite element solvers for the generalized Stokes problem. Int. J. Numer. Methods Fluids *8* (1988), 869–895. zbl MR doi
[14] *A. Greenbaum, V. Pták, Z. Strakoš*: Any nonincreasing convergence curve is possible for GMRES. SIAM J. Matrix Anal. Appl. *17* (1996), 465–469. zbl MR doi
[15] *J. L. Lions*: Some Aspects of the Optimal Control of Distributed Parameter Systems. CBMS-NSF Regional Conference Series in Applied Mathematics 6, Society for Industrial and Applied Mathematics, Philadelphia, 1972. zbl MR doi
[16] *C. C. Paige, M. A. Saunders*: Solution of sparse indefinite systems of linear equations. SIAM J. Numer. Anal. *12* (1975), 617–629. zbl MR doi
[17] *Y. Saad*: A flexible inner-outer preconditioned GMRES-algorithm. SIAM J. Sci. Comp. *14* (1993), 461–469. zbl MR doi

Author's address: *Owe Axelsson*, Institute of Geonics of the Czech Academy of Sciences, Studentská 1768, 708 00 Ostrava, Czech Republic, e-mail: `owe.axelsson@it.uu.se`.